1 Распределение Дирихле.

Смотрим на Википедии: http://en.wikipedia.org/wiki/Dirichlet_distribution.

Если случайная величина имеет распределение Дирихле с вектором параметров $\boldsymbol{\alpha}$, то вот ее мода и среднее:

$$\boldsymbol{x} \sim \mathrm{Dir}(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} \in \mathbb{R}^K \tag{1}$$

$$x_i^{\mathrm{mode}} = \frac{\alpha_i - 1}{\sum_j \alpha_j - K},\ \alpha_i > 1 \tag{2}$$

$$x_i^{\mathrm{mean}} = \frac{\alpha_i}{\sum_j \alpha_j} \tag{3}$$

В случае, если $\alpha_i \le 1$, то с модой распределения Дирихле получается хитрая вещь. Пояснение можно почитать вот здесь: http://www.quora.com/Why-can-hyper-parameters-in-beta-and-dirichlet-distribution-be

Посмотрим на плотнсть распределения Дирихле:

$$p(\boldsymbol{x}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \tag{4}$$

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{K} \alpha_i)} \tag{5}$$

Если $\alpha_i = 1$, то в плотность $p(\boldsymbol{x})$ вероятность i-того события x_i не входит: $x_i^{\alpha_i - 1} = 1$. Распределению равномерно, какое значение будет принимать x_i, поэтому мода по этой координате не определена.

Если $0 < \alpha_i < 1$, то при фиксированных всех вероятностях событий, кроме x_i, плотность зависит следующим образом:

$$p(\boldsymbol{x}) \propto \frac{1}{x_i^{1-\alpha_i}} \tag{6}$$

Это значит, что чем ближе вероятность i-того события к 0 — тем вероятнее такое распределение с точки зрения Дирихле. Но формально моды не существует, поскольку в $x_i = 0$ эта функция плотности не определена.

Если представить, что функция плотности в точке $x_i = 0$ равна $+\infty$ и это не мешает плотности интегрироваться в 1, то мода вполне себе существует и легко показать, что в таком случае:

$$x_i^{\mathrm{mode}} \propto \max\left(0, \alpha_i - 1\right). \tag{7}$$

2 Формулы для перевода модели LDA в Φ, Θ

$$\phi_t \sim \mathrm{Dir}(\lambda_t) \quad \Rightarrow \quad \begin{cases} \phi_{w,t}^{\mathrm{map}} \propto (\lambda_{w,t} - 1)_+ \\ \phi_{w,t}^{\mathrm{mean}} \propto \lambda_{w,t} \end{cases}$$

$$\theta_d \sim \mathrm{Dir}(\gamma_d) \quad \Rightarrow \quad \begin{cases} \theta_{d,t}^{\mathrm{map}} \propto (\gamma_{d,t} - 1)_+ \\ \theta_{d,t}^{\mathrm{mean}} \propto \gamma_{d,t} \end{cases}$$